NWG/RFC# 686 BH 11-MAY-75 21:41 32481
Leaving Well Enough Alone

Brian Harvey
SU-AI
Re: File Transfer Protocol
References: 354, 385, 630, 542, 640. May 10, 1975 1

Leaving Well Enough Alone 1a

I recently decided it was time for an overhaul of our FTP user and server programs. This was my first venture into the world of network protocols, and I soon discovered that there was a lot we were doing wrong--and a few things that everyone seemed to be doing differently from each other. When I enquired about this, the response from some quarters was "Oh, you're running Version 1!" 2

Since, as far as I can tell, all but one network host are running version 1, and basically transferring files OK, it seems to me that the existence on paper of an unused protocol should not stand in the way of maintaining the current one unless there is a good reason to believe that the new one is either imminent or strongly superior or both. (I understand, by the way, that FTP-2 represents a lot of thought and effort by several people who are greater network experts than I, and that it isn't nice of me to propose junking all that work, and I hereby apologize for it.) Let me list what strike me as the main differences in FTP-2 and examine their potential impact on the world. 3

1. FTP-2 uses TELNET-2. The main advantage of the new Telnet protocol is that it allows flexible negotiation about things like echoing. But the communicators in the case of FTP are computer programs, not people, and don't want any echoing anyway. The argument that new hosts might not know about old Telnet seems an unlikely one for quite some time to come; if TELNET-2 ever does really take over the world, FTP-1 could be implemented in it. 3a

2. FTP-2 straightens out the "print file" mess. This is more of a mess on paper than in practice, I think. Although the protocol document is confusing on the subject, I think it is perfectly obvious what to do: if the user specifies, and the server accepts, TYPE P (ASCII print file) or TYPE F (EBCDIC print file), then the data sent over the network should contain Fortran control characters. That is, the source file should contain Fortran controls, and should be sent over the net as is, and reformatted if necessary not by the SERVER as the protocol says but by the RECIPIENT (server for STOR, user for RETR). As a non-Fortran-user I may be missing something here but I don't think so; it is just like the well-understood TYPE E in which the data is sent in EBCDIC and the recipient can format it for local use as desired. One never reformats a file from ASCII to EBCDIC at the sending end. Perhaps the confusion happened because the protocol authors had in mind using these types to send files

directly to a line printer at the server end, and indeed maybe that's all it's good for and nobody's user program will implement TYPE P RETR. In any event, using a two-dimensional scheme to specify the combinations of ASCII/EBCDIC and ASA/normal conveys no more information than the present A-P-E-F scheme. If there is any straightening-out in FTP-2, it could only be in the handling of these files once the negotiation is settled, not in the negotiation process. 3b

3. FTP-2 approves of the Network Virtual File System concept even though it doesn't actually implement it. It seems to me that the NVFS notion is full of pitfalls, the least of which is the problem of incompatibilities in filename syntax. (For example, one would like to be able to do random access over the network, which requires that different systems find a way to accomodate each other's rules about record sizes and so on.) In any case, FTP-2 doesn't really use NVFS and I mention it here only because RFC 542 does. 3c

4. FTP-2 reshuffles reply codes somewhat. The reply codes in the original FTP-2 document, RFC 542, don't address what I see as the real reply code problems. The increased specificity of reply codes doesn't seem to be much of a virtue; if, say, a rename operation fails, it is the human user, not the FTP user program, who needs to know that it was because of a name conflict rather than some other file system error. I am all for putting such information in the text part of FTP replies. Some real problems are actually addressed in the reply code revision of RFC 640, in which the basic scheme for assigning reply code numbers is more rational than either the FTP-1 scheme or the original FTP-2 scheme. However, I think that most of the benefits of RFC 640 can be obtained in a way which does not require cataclysmic reprogramming. More on this below. 3d

5. FTP-2 was established by a duly constituted ARPAnet committee and we are duty-bound to implement it. I don't suppose anyone would actually put it that baldly, but I've heard things which amounted to that. It's silly. 3e

6. FTP-2 specifies default sockets for the data connection. Most places use the default sockets already anyway, and it is easy enough to ignore the 255 message if you want to. This is a security issue, of course, and I'm afraid that I can't work up much excitement about helping the CIA keep track of what anti-war demonstrations I attended in 1968 and which Vietnamese hamlets to bomb for the greatest strategic effect even if they do pay my salary indirectly. I could rave about this subject for pages, and probably will if I ever get around to writing an argument against MAIL-2, but for now let me just get one anecdote off my

chest: I have access to an account at an ARPAnet host because I am responsible at my own site for local maintenance of a program which was written by, and is maintained by, someone at the other site. However, the other site doesn't really trust us outsiders (the account is shared by people in my position at several other hosts) to protect their vital system security, so every week they run a computer program to generate a new random password for the account (last week's was HRHPUK) and notify us all by network mail. Well, on my system and at least one of the others, that mail isn't read protected. I delete my mail when I read it, but since it is hard enough remembering HRHPUK without them changing it every week, I naturally write it in a file on our system. That file could in principle be read protected but it isn't, since sometimes I'm in someone else's office when I want to use it, and the other passwords in it are for open guest accounts which are widely known. Moral #1: Security freaks are pretty weird. Moral #2: If you have a secret don't keep it on the ARPAnet. (In the past week I have heard about two newly discovered holes in TENEX security.) 3f

7. FTP-2 is available online and FTP-1 isn't, so new hosts can't find out how to do it. Aargh!!! What a reason for doing anything! Surely it would be less costly for someone to type it in again than for everyone to reprogram. Meanwhile these new hosts can ask Jon or Geoff or Bobby or even me for help in getting FTP up. 3g

8. FTP-2 has some changes to the strange MODEs and STRUs. This is another thing I can't get too excited about. We support only MODE S and STRU F and that will probably still be true even if we are forced into FTP-2. If the relatively few people who do very large file transfers need to improve the restart capability, they can do so within FTP-1 without impacting the rest of us. The recent implementation of paged file transfers by TENEX shows that problems of individual systems can be solved within the FTP-1 framework. If the IBM people have some problem about record structure in FTP-1, for example, let them solve it in FTP-1, and whatever the solution is, nobody who isn't affected has to reprogram. 3h

Well, to sum up, I am pretty happy with the success I've had transferring files around the network the way things are. When I do run into trouble it's generally because some particular host hasn't implemented some particular feature of FTP-1, and there's no reason to suppose they'll do it any faster if they also have to convert to FTP-2 at the same time. The main thing about FTP-2, as I said at the beginning, is that its existence is an excuse for not solving problems in FTP-1. Some such problems are quite trivial except for the fact that people are reluctant to go against anything in the

protocol document, as if the latter were the Holy Writ. A few actually require some coordinated effort. Here is my problem list: 4

1. It is almost true that an FTP user program can understand reply codes by the following simple algorithm: 4a

   a. Replies starting with 0 or 1 should be typed out and otherwise ignored. 4a1

   b. Replies starting with 2 indicate success (of this step or of the whole operation, depending on the command). 4a2

   c. Replies starting with 4 or 5 indicate failure of the command. 4a3

   d. Replies starting with 3 are only recognized in three cases: the initial 300 message, the 330 password request, and the 350 MAIL response. (Note that the user program need not distinguish which 300 message it got, merely whether or not it is expecting one right now.) 4a4

The only real problem with this, aside from bugs in a few servers whose maintainers tell me they're working on it, is the HELP command, which is not in the original protocol and which returns 0xx, 1xx, or 2xx depending on the server. (Sometimes more than one message is returned.) The word from one network protocol expert at BBN is that (a) 050 or 030 is the correct response to HELP, and (b) there is a perfectly good mechanism in the protocol for multi-line responses. Unfortunately this does not do much good in dealing with reality. There seems to be a uniform, albeit contra-protocol, procedure for handling the STAT command: 4b

```
151 information
151 information
151 ...
151 information
200 END OF STATUS                                   4b1
```

which fits right in with the above algorithm. This is despite the fact that 1xx is supposed to constitute a positive response to a command like STAT, so that according to the protocol it ought to be 4c

```
151-information
information
...
151 information                                     4c1
```

instead. (It seems to me, by the way, that 050 and 030 aren't

```
      IF1 TUTORIAL
      ...
      TUTORIAL
      IF1-TUTORIAL
```

good enough as responses to HELP since they "constitute neither a positive nor a negative acknowledgement" of the HELP command and thus don't tell the user program when it ought to ask the human user what to do next.) I suggest that despite the protocol, a 200 response be given by all servers at the end of whatever other HELP it gives as of, let's say, June 1. The alternatives are either to let the current rather chaotic situation continue forever while waiting for FTP-2, or to try to standardize everyone on a multi-line 1xx for both HELP and STAT. I'm against changing STAT, which works perfectly for everyone as far as I can tell, and it should be clear that I'm against waiting for FTP-2. Unfortunately there is no real mechanism for "officially" adopting my plan, but I bet if TENEX does it on June 1 the rest of the world will come along. 4d

2. Another reply code problem is the use of 9xx for "experimental" replies not in the protocol. This includes the BBN mail-forwarding message and one other that I know of. This procedure is sanctioned by RFC 385, but it seems like a bad idea to me. For one thing, the user program has no way of knowing whether the reply is positive, negative, or irrelevant. The examples I've been burned by all should have been 0xx messages. I propose that all such messages be given codes in the 000-599 range, chosen to fit the scheme given above for interpreting reply codes. x9x or xx9 could be used to indicate experiments. 4e

3. One more on reply codes: RFC 630 (the one about the TENEX mod to the reply codes for MAIL and MLFL) raises the issue of "temporary" versus "permanent" failures within the 4xx category. RFC 640 deals with this question in the FTP-2 context by changing the meaning of 4xx and 5xx so that the former are for temporary errors and the latter are for permanent errors. I like this idea, and I think it could easily be adapted for FTP-1 use in a way which would allow people to ignore the change and still win. At present, I believe that the only program which attempts to distinguish between temporary and permanent errors is the TENEX mailer. For other programs, no distinction is currently made between 4xx and 5xx responses; both indicate failure, and any retrials are done by the human user based on the text part of the message. A specific set of changes to the reply codes is proposed below. 4f

Perhaps I should make a few more points about RFC 640, since it's the best thing about FTP-2 and the only argument for it I find at all convincing. Let me try to pick out the virtues of 640 and indicate how they might be achieved in FTP-1. 4g

a. The 3xx category is used uniformly for "positive intermediate replies" where further negotiation in the Telnet

connection is required, as for RNFR. I'm afraid this one can't be changed without affecting existing user programs. (One of my goals here is to enable existing user programs to work while some servers continue as now and others adopt the suggestions I make below.) However, although this 3xx idea is logically pleasing, it is not really necessary for a simple-minded user program to be able to interpret replies. The only really new 3xx in RFC 640 is the 350 code for RNFR. But this would only be a real improvement for the user program if there were also a 2xx code which might be returned after RNFR, which is not the case. 640 also abolishes the 300 initial connection message with 220, but again there is clearly no conflict here. 4g1

b. The use of 1xx is expanded to include what is now the 250 code for the beginning of a file transfer. The idea is that a 1xx message doesn't affect the state of the user process, but this is not really true. Consider the file transfer commands. The state diagram on page 13 of RFC 640 is slightly misleading. It appears as if 1xx replies are simply ignored by the user program. In reality, that little loop hides a lot of work: the file transfer itself! If the server replied to the file transfer command immediately with a 2xx message, it would be a bug in the server, not a successful transfer. The real state diagram is more like 4g2

```
B --> cmd --> W --> 1 --> W --> 2 --> S
```

(with branches out from the "W"s for bad replies). It should be clear from this diagram that the user program, if it trusts the server to know what it's doing, can expect a 2xx instead of the 1xx without getting confused, since it knows which of the W states it's in. In fact, the use of 1xx in file transfer is very different from its other uses, which are indeed more like the 0xx and 1xx replies in FTP-1. I'd call this particular point a bug in RFC 640. 4g3

c. Automatic programs which use FTP (like mailers) can decide whether to queue or abandon an unsuccessful transfer based on the distinction between 4xx and 5xx codes. I like this idea, although those temporary errors virtually never happen in real life. This could be accomplished in FTP-1 by moving many of the 4xx replies to 5xx. Mailers would be modified to use the first digit to decide whether or not to retry. This scheme does not cause any catastrophes if some server is slow in converting; it merely leads to unnecessary retries. A few CPU cycles would be wasted in the month following the official switch. Thus, this feature is very different from (a) and (b), which could lead to catastrophic failures if not

implemented all at once. (Yes, I know that FTP-2 is supposed to be done on a different ICP socket. I am not discussing FTP-2 but whether its virtues can be transferred to FTP-1.) The specific codes involved are listed below. 4g4

d. The use of the second digit to indicate the type of message. (The proposed division is not totally clean; for example, why is 150 ("file status okay; about to open data connection") considered to be more about the file system than about the data connection?) This can easily be done, since the second digit is not currently important to any user process--the TENEX mailer is, in this plan, already due for modification because of (c). Since this is mostly an aesthetic point, I'm hesitant to do it if it would be difficult for anyone. In particular, I would want to leave the 25x messages alone, in case some user programs distinguish these. This is especially likely for the ones which are entirely meant for the program: 251 and 255. Therefore I propose that if this idea is adopted in FTP-1 the meanings of x2x and x5x be interchanged. This proposal is reflected in the specific list below. 4g5

4. The print file thing again. Let's get it made "official" that it is the recipient, not the server, who is responsible for any reformatting which is to be done on these files. After all, the recipient knows what his own print programs want. 4h

Let me summarize the specific changes to FTP-1 I'd like to see made, most of which are merely documentation changes to reflect reality: 5

1. HELP should return 200. All commands should return 2xx if successful, and I believe all do except HELP. 5a

2. The definition of 1xx messages should be changed to read: "Informative replies to status inquiries. These constitute neither a positive nor a negative acknowledgment." 5b

3. Experimental reply codes should be of the form x9x or xx9, where the first digit is chosen to reflect the significance of the reply to automated user programs. Reply codes greater than 599 are not permitted. The xx9 form should be used if the reply falls into one of the existing categories for the second digit. User programs are encouraged to determine the significance of the reply from the first digit, rather than requiring a specific reply code, when possible. 5c

4. The STAT command with no argument is considered a request for a directory listing for the current working directory, except that it may be given along with TELNET SYNCH while a transfer is

in progress, in which case it is a request for the status of that transfer. (Everyone seems to do the first part of this. I'm not sure if anyone actually implements the second. This is just getting the protocol to agree with reality.) The reply to a STAT command should be zero or more 1xx messages followed by a 200. 5d

5. TYPEs P and F mean that the source file contains ASA control characters and that the recipient program should reformat it if necessary. 5e

Here is a list of the current FTP-1 replies, and how they should be renumbered for the new scheme. The changes from 4xx to 5xx should be REQUIRED as of June 1; changes in the second or third digit are not so important. (As explained above, it will not be catastrophic even if some hosts do not meet the requirement.) The list also contains one new possible reply adapted from RFC 640. 5f

OLD	NEW	TEXT
0x0	0x0	(These messages are not very well defined nor very important. Servers should use their judgment.)
100	110	System status reply. (Since nobody does STAT as in the protocol, this may be a moot point.)
150	150	"File status reply." (If this were really that, it would be switched to 120, but I believe what is meant is the response to a bare STAT in mid-transfer, which is more a connection status reply than a file status reply.
151	121	Directory listing reply.
200	200	Last command ok.
201	251	ABOR ok.
202	252	ABOR ignored, no transfer in progress.
new	206	Command ignored, superfluous here.
230	230	Login complete.
231	231	Logout complete.
232	232	Logout command will be processed when transfer is complete.
250	250	Transfer started correctly.
251	251	MARK yyyy = mmmm
252	252	Transfer completed ok.
253	223	Rename ok.
254	224	Delete ok.
255	255	SOCK nnnn
256	256	Mail completed ok.
300	300	Connection greeting
301	301	Command incomplete (no crlf)
330	330	Enter password
350	350	Enter mail.
400	huh?	"This service not implemented." I don't understand this; how does it differ from 506? If it means no FTP at all, who gave the message? Flush.

```
```

401	451	Service not accepting users now, goodbye.
430	430	Foo, you are a password hacker!
431	531	Invalid user or password.
432	532	User invalid for this service.
434	454	Logout by operator.
435	455	Logout by system.
436	456	Service shutting down.
450	520	File not found.
451	521	Access denied.
452	452	Transfer incomplete, connection closed.
453	423	Transfer incomplete, insufficient storage space.
454	454	Can't connect to your socket.
500	500	Command gibberish.
501	501	Argument gibberish.
502	502	Argument missing.
503	503	Arguments conflict.
504	504	You can't get there from here.
505	505	Command conflicts with previous command.
506	506	Action not implemented.